

What is claimed is:

1. A document image processor comprising image inputting means preparing document images by reading a paper document, region dividing means dividing the document image into a plurality of regions, and title-region extracting means extracting title regions from the entire regions according to a region average character size equivalent to an average size of characters that is calculated per region divided by the region dividing means,

wherein the title-region extracting means compares each region average character size and an extracting criterion respectively; the extracting criterion that is a total average character size multiplied by an extracting parameter; the total average character size calculated as a value equivalent to an average size of all characters included in the entire regions; and extracts as a title region regions with the region average character size larger than the extracting criterion.

2. A document image processor according to claim 1, wherein the title-region extracting means calculates the region average character size and the total average character size based on an average height of characters.

3. A document image processor according to claim 1, wherein the title-region extracting means calculates the region average character size and the total average character size based on an average width of characters.

4. A document image processor according to claim 1, wherein the title-region extracting means calculates the region average character size and the total average character size based on an average area of characters.

5. A document image processor according to claim 1, wherein the title-region extracting means calculates the extracting criteria on a plurality of levels by using the extracting parameters on a plurality of levels.

6. A document image processor according to claim 1, wherein the title-region extracting means calculates the extracting criteria on a plurality of levels by using the extracting parameters on a plurality of levels and extracts each title region corresponding to each level attribute indicating the level of the extracting.

7. A document image processor according to claim 2 or 3, wherein the title-region extracting means determines the extracting parameters on a plurality of levels based on a maximum value of the region average character size divided by the total average character size.

8. A document image processor according to claim 1, wherein the title-region extracting means adopts the trim average method for calculating the total average character size and the region average

character size according to characters excluding both characters larger than the specific ratio and characters smaller than the specific ratio.

9. A document image processor according to claim 1, which
5 comprising correcting means correcting character strings of the extracted title regions.

10. A document image processor according to claim 1, wherein the document image is configured by a plurality of pages.

11. A document title extracting method of a document image processor comprising:

inputting and preparing document images by reading a paper document;

15 dividing a plurality of regions from a document image;

calculating a region average character size equivalent to the average size of characters per region; and

extracting title region from the entire regions based on the region average character size,

20 in which the step of calculating comprises calculating a total average character size equivalent to the average size of characters in the entire regions,

and further comprising comparing the region average character size and a extracting criterion that is the total average character size
25 multiplied by an extracting parameter; and

AB
on

in which the step of extracting title region comprises extracting as a title region regions with the region average character size larger than the extracting criterion.

5 12. A document title extracting method of a document image processor according to claim 11, in which the step of calculating comprises calculating the region average character size and the total average character size based on an average height of characters.

10 13. A document title extracting method of a document image processor according to claim 11, in which the step of calculating comprises calculating the region average character size and the total average character size based on an average width of characters.

15 14. A document title extracting method of a document image processor according to claim 11, in which the step of calculating comprises calculating the region average character size and the total average character size based on an average area of characters.

20 15. A document title extracting method of a document image processor according to claim 14, in which the step of extracting titles comprises calculating the extracting criteria on a plurality of levels by using the extracting parameters on a plurality of levels.

25 16. A document title extracting method of a document image

processor according to claim 14, in which the step of extracting titles comprises calculating the extracting criterions on a plurality of levels by using the extracting parameters on a plurality of levels and extracting each title region corresponding to each level attribute indicating the level of the extracting.

17. A document title extracting method of a document image processor according to claim 15 or 16, in which the step of extracting titles comprise determining the extracting parameters on a plurality of levels based on a maximum value of the region average character size divided by the total average character size.

18. A document title extracting method of a document image processor according to claim 11, in which the step of extracting title comprises calculating the total average character size and the region average character size according to the trim average method that calculates the average of characters excluding both the characters larger than the specific ratio and the characters smaller than the specific ratio.

19. A document title extracting method of a document image processor according to claim 11, further comprising the step of:

correcting character strings of the extracted title regions.

20. A document title extracting method of a document image

processor according to claim 11, wherein the document image is configured by a plurality of pages.

21. A recording medium for recording programs comprising:
dividing document images prepared by reading a paper document into a plurality of regions;

calculating per region a region average character size equivalent to an average size of characters in a region and a total average character size equivalent to an average size of characters in the entire regions;

comparing each region average character size and extracting criterion that is the total average character size multiplied by the extracting parameter; and

extracting regions with the region average character size larger than the extracting criterion as a title region.

22. A document image processor comprising image inputting means preparing document images by reading a paper documents and storage means storing the document images, further comprising:

reference tag information storage means storing reference tag information together with each attribute value of the reference tag information;

mark extracting means extracting specific marks imparted to the paper document by a user;

calculating means calculating characteristics values

representing respective characteristics of the marks according to the variance of pixels composing the specific mark; and

document tag information imparting means selecting a specific reference tag information according to the attribute value and the characteristics value and imparting the specific reference tag information to the document image.

23. A document image processor according to claim 22, wherein the mark extracting mean extracts specific marks on a specific sheet.

24. A document image processor according to claim 23, wherein the mark extracting means recognizes a sheet attached with a two-dimensional code as the specific sheet.

25. A document image processor according to claim 22, wherein the mark extracting means extracts the specific marks on the blank of the paper document.

26. A document image processor according to claim 22, wherein the paper document is configured by a plurality of pages.

27. A document tag information imparting method of a document image processor imparting the document tag information of the document image prepared by reading a paper document, which comprising:

storing reference tag information together with each attribute

value of the reference tag information;

extracting specific marks imparted to the paper document by a user;

calculating numerical values representing respective characteristics of the marks in accordance with the variance of the pixels composing the specific mark; and

imparting to the document image the specific reference tag information selected according to the attribute value and the characteristics value.

28. A document tag information imparting method of a document image processor according to claim 27, wherein the paper document is configured by a plurality of pages.

29. A recording medium recording programs comprising:
extracting the specific marks imparted to a paper document by a user when document images are prepared by reading the paper document;

calculating numerical values representing respective characteristics of the marks based on the variance of the pixels of the mark; and

selecting document tag information to be attached to the document image from the nominee of the document tag information based on the numerical values.